

Plan Overview

A Data Management Plan created using DMPTool

Title: Data Transfer Service

Creator: Elisha Wood-charlson - **ORCID:** [0000-0001-9557-7715](https://orcid.org/0000-0001-9557-7715)

Affiliation: Lawrence Berkeley National Laboratory (lbl.gov)

Principal Investigator: Elisha Wood-Charlson

Project Administrator: AJ Ireland

Contributor: Kjiersten Fagnan, Jeffrey N Johnson

Funder: United States Department of Energy (DOE) (energy.gov)

Template: Department of Energy (DOE): Office of Science (Updated 2022)

Project abstract:

The transfer of data between BER funded programs often requires researchers to manually download and upload files. This takes time, adds the potential for error (both human and data integrity), and often removes any citation/credit information for the data being transferred. We will build a data transfer service that can be used by any BER funded program to directly request or send data objects from other programs that is accompanied by appropriate provenance and citation metadata. By ensuring data object integrity and retention of credit metadata, DTS will enable better tracking of data reuse across the BER funded portfolio. Finally, the DTS aims to reduce the time and effort required by researchers to accomplish their science.

Start date: 10-01-2023

End date: 09-30-2025

Last modified: 06-15-2023

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Data Transfer Service

Data Types and Sources

A brief, high-level description of the data to be generated or used through the course of the proposed research and which of these are considered [digital research data](#) necessary to [validate](#) the research findings

The Data Transfer Service will not generate any new data. The goal of the DTS is to more efficiently transfer existing data between data platforms, ensuring data integrity and credit metadata associated with the data object are intact and travel with the data itself.

DTS will connect BER data platforms, and any trusted non-BER data platform, that is interested in connecting to the service.

Content and Format

A statement of plans for data and metadata content and format including, where applicable, a description of documentation plans, annotation of relevant software, and the rationale for the selection of appropriate standards. (Existing, accepted community standards should be used where possible. Where community standards are missing or inadequate, the DMP could propose alternate strategies that facilitate sharing, and should advise the sponsoring program of any need to develop or generalize standards.)

All software and data models developed for the DTS will be open source, and made available with robust documentat on GitHub. Documentation includes how to use the API, detailed information around fields in the credit metadata schema, and comprehensive tutorials on how to connect to the service via each participating data platform.

Sharing and Preservation

The anticipated means for sharing and the rationale for any restrictions on who may access the data and under what conditions

All code and transfer logs (source/destination, data identifier, md5) will be made available via the GitHub repository.

A timeline for sharing and preservation that addresses both the minimum length of time the

data will be available and any anticipated delay to data access after research findings are published.

Question not answered.

Any special requirements for data sharing, for example, proprietary software needed to access or interpret data, applicable policies, provisions, and licenses for re-use and re-distribution, and for the production of derivatives, including guidance for how data and data products should be cited.

Any resources and capabilities (equipment, connections, systems, software, expertise, etc.) requested in the research proposal that are needed to meet the stated goals for sharing and preservation. (This could reference the relevant section of the associated research proposal and budget request).

DTS has requested funding to ensure that the production service is deployed on hardware that serves the JGI and KBase programs, in the Integrated Genomics Building on the Berkeley Lab campus.

Cost/benefit considerations to support whether/where the data will be preserved after direct project funding ends and any plans for the transfer of responsibilities for sharing and preservation.

Question not answered.

Whether, when, or under what conditions the management responsibility for the research data will be transferred to a third party (e.g. institutional, or community repository).

After the completion of this project, maintenance of the DTS will be performed by JGI and KBase team members involved in ongoing JGI-KBase co-development efforts.

Any other future decision points regarding the management of the research data including plans to reevaluate the costs and benefits of data sharing and preservation?

Question not answered.

Protection

A statement of plans, where appropriate and necessary, to protect confidentiality, personal privacy, [Personally Identifiable Information](#), and U.S. national, homeland, and economic security; recognize proprietary interests, business confidential information, and intellectual property rights; and avoid significant negative impact on innovation, and U.S.

competitiveness.

No PII will be stored by the DTS. Authentication protocols will be used to ensure transfer of data between portals (e.g., ORCID). Once the transfer is complete, authentication information is not retained. Transfer logs are only to track the location of data across the portals, to avoid duplicate transfers and to quantify movement of data between programs.

Rationale

A discussion of the rationale or justification for the proposed data management plan including, for example, the potential impact of the data within the immediate field and in other fields, and any broader societal impact.

The DTS is not generating new data, but will greatly improve the researcher experience in needing to move data from one or more portals. The aim is to simplifying curation efforts, streamlining the steps between data production and publication, and directly support FAIR and open data best practices.

Planned Research Outputs

Software - "Data Transfer Service"

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Data Transfer Service	Software	Unspecified	Open	None specified		Creative Commons Zero v1.0 Universal	None specified	No	No